INFORMATION AND CONTROL 24, 279–291 (1974)

On Measure Theoretical Structure of Multivariate Information Among Random Variables

TAKEO YUKIMACHI

Department of Administration Engineering, Faculty of Engineering, Keio University, 832 Hiyoshi-cho, Kohoku-ku, Yokohama 223, Japan

This paper is concerned with a quantity called "multivariate information" among random variables. By defining it as a specialization of the information proposed by Kullback and Leibler, we will discuss fundamental properties of it, which are direct applications of the results of mathematical information theory.

We will derive some formulas about multivariate information closely related to the absolute continuity of the joint probability distribution with respect to some other distribution.

1. INTRODUCTION

As is generally known, the Kullback–Leibler information number is given by the form

$$I(2:1) = \int \log \frac{dP_2}{dP_1}\, dP_2$$

(Kullback and Leibler, 1951). This information number may be one of the most generalized forms among all the known information numbers. As the form of $I(2{:}\ 1)$ is too broad, some appropriate restriction for P_1 and P_2 are often required if we want to represent the characteristics of any information system. The mutual information $I(\xi, \eta)$ between the random variables ξ and η is a special form of $I(2{:}\ 1)$ which is defined by restricting $P_2 = P_{\xi\eta}$ and $P_1 = P_\xi P_\eta$, where $P_{\xi\eta}$ is the joint probability distribution of ξ and η, and P_ξ and P_η are the probability distributions of ξ and η, respectively. The mutual information plays very important roles in information theory, and the measure theoretical properties of it have been precisely studied by many investigators (Kolmogorov, Yaglom, and Gelfand, 1956; Perez, 1959; Dobrushin, 1959; and Pinsker, 1960).

However, the mutual information, which depends essentially upon two random variables in the defining form, seems to be not sufficient if we shall

643/24/3-6

be concerned with any information system which has many components such as a network system. The "multivariate information" which we shall define in this paper is a generalization of the mutual information, and it is also a special case of the Kullback–Leibler information number. As will be described later, the multivariate information is a quantity to represent the degree of the statistical interrelation among finitely many random variables, $\xi_1, \ldots, \xi_n$. In the case in which the random variables have finite schemes or a joint probability density function with respect to the Lebesgue measure, Watanabe already introduced such a quantity of information by the form $\sum_{i=1}^{n} H(\xi_i) - H(\xi_1, \ldots, \xi_n)$, where H means the entropy function (Watanabe, 1960, 1969, etc.) Our intention here is to provide an expression of such an information in a more generalized form and to investigate its properties on the basis of measure theory.

2. Notation and Preliminaries

Suppose ξ_i, $i = 1, \ldots, n$, are random variables defined on the basic space (Ω, B_ω, P), taking values in measurable spaces (X_i, B_{X_i}), $i = 1, \ldots, n$, respectively. We denote by ξ_1^n the n-dimensional random variable $(\xi_1, \ldots, \xi_n)$ which takes values in $(X_1^n, B_{X_1^n}) = (\times_{i=1}^{n} X_i, B_{\times_{i=1}^{n} X_i})$, where $B_{X_1^n} = B_{\times_{i=1}^{n} X_i}$ is the σ-algebra generated by all sets of the form $E_1 \times \cdots \times E_n$, $E_i \in B_{X_i}$.

Let $P_{\xi_1}, \ldots, P_{\xi_n}$ and $P_{\xi_1^n}$ be the probability measures on $(X_1, B_{X_1}), \ldots, (X_n, B_{X_n})$ and $(X_1^n, B_{X_1^n})$, respectively. Let $Q_{\xi_1^n}$ denote another probability measure on $(X_1^n, B_{X_1^n})$ such that

$$Q_{\xi_1^n}\left(\mathop{\times}_{i=1}^{n} E_i\right) = \prod_{i=1}^{n} P_{\xi_i}(E_i),$$

for every $E_i \in B_{X_i}$, $i = 1, \ldots, n$.

We shall frequently use the notations $P_{\xi_1 \cdots \xi_n}$ and $Q_{\xi_1 \cdots \xi_n}$ in place of $P_{\xi_1^n}$ and $Q_{\xi_1^n}$ for convenience.

In this paper we assume that the conditional probabilities can be always defined as measures. Let η be a random variable taking values in (Y, B_Y) and $P_{\xi_1^n \eta} = P_{\xi_1 \cdots \xi_n \eta}$ be the joint distribution of $\xi_1, \ldots, \xi_n$ and η. The conditional probabilities are denoted by $P_{\xi_1^n | \eta}$, $P_{\xi_i | \eta}$, $i = 1, \ldots, n$. And we define a set function $Q_{\xi_1^n | \eta}$ by the formula

$$Q_{\xi_1^n | \eta}\left(\mathop{\times}_{i=1}^{n} E_i \times F\right) = \int_F \prod_{i=1}^{n} P_{\xi_i | \eta}(E_i \mid y)\, dP_\eta$$

for any rectangles of the form $\times_{i=1}^{n} E_i \times F$, $E_i \in B_{X_i}$, $F \in B_Y$ (cf. Dobrushin, 1959).

Let $B_{X_1^n} \times B_Y$ be the algebra of all the above rectangles and their finite sums. Then $Q_{\xi_1^n|\eta}$ is additive for any set in $B_{X_1^n} \times B_Y$, and it is easily seen that if $\{N_k\}$ is a nonincreasing sequence of sets in $B_{X_1^n} \times B_Y$ such that $\bigcap_{k=1}^{\infty} N_k = \phi$, then $\lim_{k\to\infty} Q_{\xi_1^n|\eta}(N_k) = 0$. Therefore $Q_{\xi_1^n|\eta}$ can be uniquely extended as a probability measure on $(X_1^n \times Y, B_{X_1^n \times Y})$, where $B_{X_1^n \times Y}$ is the σ-algebra generated by $B_{X_1^n} \times B_Y$.

If $P_{\xi_1^n|\eta}$ coincides with $Q_{\xi_1^n|\eta}$, we say that $\xi_1, \ldots, \xi_n$ are conditionally independent with respect to η.

If there is given any measurable space (X, B_X), then a finite system $\{E^1, \ldots, E^s\}$ such that $E^j \in B_X$, $1 \leqslant j \leqslant s < \infty$, $\bigcup_{j=1}^{s} E^j = X$ and $E^j \cap E^k = \phi$, $j \neq k$, is called a B_X-measurable partition of X and written δ_X. If there is another measurable space (Y, B_Y) and if δ_Y is a B_Y-measurable partition of Y, we denote by $\delta_X \times \delta_Y$ the refinement of δ_X and δ_Y which is one of the $B_{X\times Y}$-measurable partitions of $X \times Y$. Any $B_{X\times Y}$-measurable partition of $X \times Y$ is written $\delta_{X\times Y}$.

The family of all δ_X is denoted by Δ_X.

3. Multivariate Information

In this section we shall define the multivariate information and discuss several properties of it, most of which are the direct extensions of the corresponding properties of mutual information.

Using the previous notations, we define a function J of $\delta_{X_1^n} = \{E^1, \ldots, E^s\}$ by the formula

$$J(P_{\xi_1^n}, Q_{\xi_1^n}; \delta_{X_1^n}) = \sum_i P_{\xi_1^n}(E^i) \log(P_{\xi_1^n}(E^i))/(Q_{\xi_1^n}(E^i)).$$

The multivariate information among $\xi_1, \ldots, \xi_n$, written $I(\xi_1^n)$ or $I(\xi_1, \ldots, \xi_n)$, is defined by the formula

$$I(\xi_1^n) = I(\xi_1, \ldots, \xi_n) = \sup_{\delta_{X_1^n} \in \Delta_{X_1^n}} J(P_{\xi_1^n}, Q_{\xi_1^n}; \delta_{X_1^n}).$$

Similarly, we define the conditional multivariate information, written by $I(\xi_1^n \mid \eta)$ or $I(\xi_1, \ldots, \xi_n \mid \eta)$, among $\xi_1, \ldots, \xi_n$ relative to η, by the formula

$$\begin{aligned} I(\xi_1^n \mid \eta) &= I(\xi_1, \ldots, \xi_n \mid \eta) \\ &= \sup_{\delta_{X_1^n \times Y} \in \Delta_{X_1^n \times Y}} J(P_{\xi_1^n \eta}, Q_{\xi_1^n|\eta}; \delta_{X_1^n \times Y}). \end{aligned}$$

The following proposition is immediate from the fact that each of $I(\xi_1{}^n)$ and $I(\xi_1{}^n \mid \eta)$ is a special form of the Kullback–Leibler information number (Kullback and Leibler, 1951).

PROPOSITION 1. (a) *Always we have*

$$I(\xi_1{}^n) \geqslant 0,$$

and the equality holds if and only if $\xi_1, \ldots, \xi_n$ *are independent.*

(b) *Always we have*

$$I(\xi_1{}^n \mid \eta) \geqslant 0,$$

and the equality holds if and only if $\xi_1, \ldots, \xi_n$ *are conditionally independent with respect to* η.

Roughly speaking, the quantity $I(\xi_1, \ldots, \xi_n)$ represents the degree of the statistical dependence among the random variables, $\xi_1, \ldots, \xi_n$, and the quantity $I(\xi_1, \ldots, \xi_n \mid \eta)$ means the degree of the statistical dependence which still remains among $\xi_1, \ldots, \xi_n$, after η is specified.

For example, let us consider a network system $(\xi_1, \ldots, \xi_n)$, which are any observations representing the state of the system. In this case, $I(\xi_1, \ldots, \xi_n)$ is a measure of the degree of the interrelation among the system components. As can be seen, if there exists any deterministic relation among them, $I(\xi_1, \ldots, \xi_n)$ becomes infinite, except for the case where the joint distribution of $(\xi_1, \ldots, \xi_n)$ is atomic with a finite entropy. On the contrary, if $I(\xi_1, \ldots, \xi_n) = 0$, the system may be regarded as an ideal chaos.

Now we shall give several propositions. As is easily verified, we can derive the following proposition by applying the well-known method of the proof in the Dobrushin's theorem on mutual information (Dobrushin, 1959).

PROPOSITION 2. *Let* $\times_{i=1}^{n} \Delta_{X_i}$ *be the family of all rectangular partitions of* $X_1{}^n$ *of the form* $\delta_{X_1} \times \cdots \times \delta_{X_n}$. *Then*

$$I(\xi_1{}^n) = \sup_{\times_{i=1}^n \delta_{X_i} \in \times_{i=1}^n \Delta_{X_i}} J(P_{\xi_1{}^n}, Q_{\xi_1{}^n}; \delta_{X_1} \times \cdots \times \delta_{X_n}).$$

Similarly,

$$I(\xi_1{}^n \mid \eta) = \sup_{\times_{i=1}^n \delta_{X_i} \times \delta_Y \in \times_{i=1}^n \Delta_{X_i} \times \Delta_Y} J(P_{\xi_1{}^n \eta}, Q_{\xi_1{}^n \mid \eta}; \delta_{X_1} \times \cdots \times \delta_{X_n} \times \delta_Y).$$

We also have

PROPOSITION 3. *If $\xi_1^n = (\xi_1, \ldots, \xi_n)$ and η are independent, then*

$$I(\xi_1^n \mid \eta) = I(\xi_1^n).$$

Proof. If $(\xi_1, \ldots, \xi_n)$ and η are independent, then it is easily seen that, for each i, and for every $E_i \in B_{X_i}$,

$$P_{\xi_i \mid \eta}(E_i \mid y) = P_{\xi_i}(E_i), \quad \text{a.s.}$$

Therefore

$$Q_{\xi_1^n \mid \eta}\left(\bigtimes_{i=1}^{n} E_i \times F\right) = \prod_{i=1}^{n} P_{\xi_i}(E_i)\, P_\eta(F),$$

and

$$J(P_{\xi_1^n \eta}, Q_{\xi_1^n \mid \eta}; \delta_{X_1} \times \cdots \times \delta_{X_n} \times \delta_Y) = J(P_{\xi_1^n}, Q_{\xi_1^n}; \delta_{X_1} \times \cdots \times \delta_{X_n}). \tag{1}$$

In view of (1) and Proposition 2, we have $I(\xi_1^n \mid \eta) = I(\xi_1^n)$.

Now let $\zeta_1^n = (\zeta_1, \ldots, \zeta_n)$ be a random variable taking values in

$$(Z_1^n, B_{Z_1^n}) = \left(\bigtimes_{i=1}^{n} Z_i, B_{\bigtimes_{i=1}^n Z_i}\right).$$

We shall say that ξ_1^n is subordinate to ζ_1^n if for every $E \in B_{X_1^n}$, there exists some $D \in B_{Z_1^n}$ such that

$$P(E^{-1} \Delta D^{-1}) = 0,$$

where E^{-1} and D^{-1} are the inverse images of E and D, respectively, namely, $E^{-1} = \{\omega; \xi_1^n(\omega) \in E\}$ and $D^{-1} = \{\omega; \zeta_1^n(\omega) \in D\}$.

Furthermore, ξ_1^n is called separately subordinate to ζ_1^n if for every rectangle $\bigtimes_{i=1}^n E_i$, $E_i \in B_{X_i}$, there exists some rectangle $\bigtimes_{i=1}^n D_i$, $D_i \in B_{Z_i}$, such that

$$P(E_i^{-1} \Delta D_i^{-1}) = 0$$

for each i.

As $B_{X_1^n}$ and $B_{Z_1^n}$ are respectively generated by the algebras $\bigtimes_{i=1}^n B_{X_i}$ and $\bigtimes_{i=1}^n B_{Z_i}$, it can be shown that if ξ_1^n is separately subordinate to ζ_1^n, then ξ_1^n is subordinate to ζ_1^n.

In fact, for any sequence $\{\epsilon_k\}$ such that $\epsilon_k > 0$, $\sum_{k=1}^\infty \epsilon_k < \infty$, and for any $E \in B_{X_1^n}$, we can find some sequence $\{E^k\}$, $E^k \in \bigtimes_{i=1}^n B_{X_i}$, satisfying $E^k \supset E^{k+1} \supset \cdots \supset E$, and $P(E^{k^{-1}} - E^{-1}) < \epsilon_k$.

From the definition of the separate subordination, there exists some $D^k \in \mathsf{X}_{i=1}^{n} B_{Z_i}$, such that

$$P(E^{k^{-1}} \Delta D^{k^{-1}}) = 0.$$

Letting $D = \bigcap_{m=1}^{\infty} \bigcup_{k \geqslant m} D^k$, we have

$$\begin{aligned} P(D^{-1} - E^{-1}) &= P\left(\bigcap_{m=1}^{\infty}\left(\bigcup_{k\geqslant m} D^{k^{-1}} - E^{-1}\right)\right) \\ &\leqslant \lim_{m\to\infty}\left\{P\left(\bigcup_{k\geqslant m} D^{k^{-1}} - \bigcup_{k\geqslant m} E^{k^{-1}}\right) + P\left(\bigcup_{k\geqslant m} E^{k^{-1}} - E^{-1}\right)\right\} \\ &= \lim_{m\to\infty}\left\{\sum_{k\geqslant m} P(D^{k^{-1}} - E^{k^{-1}}) + \sum_{k\geqslant m} P(E^{k^{-1}} - E^{-1})\right\} \\ &\leqslant \lim_{m\to\infty} \sum_{k\geqslant m} \epsilon_k . \end{aligned}$$

On the other hand, we have

$$\begin{aligned} P(E^{-1} - D^{-1}) &= P\left(\bigcup_{m=1}^{\infty} \bigcap_{k\geqslant m} (E^{-1} - D^{k^{-1}})\right) \\ &\leqslant \sum_{m=1}^{\infty} P\left(\bigcap_{k\geqslant m} (E^{-1} - D^{k^{-1}})\right) = 0, \end{aligned}$$

because $E \subset E^k$ for every k and $P(E^{-1} - D^{k^{-1}}) = 0$. As $\sum_{k=1}^{\infty} \epsilon_k < \infty$, it follows that

$$P(E^{-1} \Delta D^{-1}) = 0.$$

PROPOSITION 4. *If ξ_1^n is separately subordinate to ζ_1^n, then*

$$I(\xi_1^n \mid \eta) \leqslant I(\zeta_1^n \mid \eta). \tag{2}$$

In particular,

$$I(\xi_1^n) \leqslant I(\zeta_1^n). \tag{3}$$

Proof. From the assumption, there exists $D_i \in B_{Z_i}$, for every $E_i \in B_{X_i}$, such that

$$P(E_i^{-1} \Delta D_i^{-1}) = 0. \tag{4}$$

(4) means that for any $F \in B_Y$,

$$P_{\xi_i\eta}(E_i \times F) = P_{\zeta_i\eta}(D_i \times F). \tag{5}$$

In view of the definition of the conditional probability and Eq. (5), we have

$$P_{\xi_i|\eta}(E_i \mid y) = P_{\zeta_i|\eta}(D_i \mid y) \quad \text{a.s.}$$

Therefore, it is clear that D_i, $i = 1,..., n$, satisfy

$$Q_{\xi_1^n|\eta}\left(\bigtimes_{i=1}^{n} E_i \times F\right) = Q_{\zeta_1^n|\eta}\left(\bigtimes_{i=1}^{n} D_i \times F\right). \tag{6}$$

Now suppose that $\delta_{X_i} = \{E_i^1,..., E_i^{s_i}\}$, $i = 1,..., n$, and $\delta_Y = \{F^1,..., F^s\}$ are partitions of X_i, $i = 1,..., n$, and Y, respectively, and $D_i^1,..., D_i^{s_i}$ are the corresponding sets in B_{Z_i} satisfying $P(E_i^{j_i^{-1}} \varDelta D_i^{j_i^{-1}}) = 0$ for each $j_i = 1,..., s_i$, $i = 1,..., n$.

As the system $\{D_i^1,..., D_i^{s_i}\}$ is not always a partition of Z_i, we set

$$C_i^{j_i} = D_i^{j_i} - \bigcup_{k=1}^{j_i-1} D_i^k, \qquad \text{for} \quad 1 \leqslant j_i < s_i,$$

$$C_i^{s_i} = Z_i - \bigcup_{k=1}^{s_i-1} D_i^k.$$

Then the system $\delta_{Z_i} = \{C_i^1,..., C_i^{s_i}\}$ is a partition of Z_i, and it can be easily verified that

$$P(E_i^{j_i^{-1}} \varDelta C_i^{j_i^{-1}}) = 0,$$

for each $j_i = 1,..., s_i$, $i = 1,..., n$.

Using (6), we have

$$Q_{\xi_1^n|\eta}\left(\bigtimes_{i=1}^{n} E_i^{j_i} \times F^j\right) = Q_{\zeta_1^n|\eta}\left(\bigtimes_{i=1}^{n} C_i^{j_i} \times F^j\right) \tag{7}$$

for each $E_i^{j_i} \in \delta_{X_i}$, $C_i^{j_i} \in \delta_{Z_i}$, $i = 1,..., n$, and $F^j \in \delta_Y$.

On the other hand, we have

$$P\left(\bigcap_{i=1}^{n} E_i^{j_i^{-1}} \varDelta \bigcap_{i=1}^{n} C_i^{j_i^{-1}}\right) \leqslant \sum_{i=1}^{n} P(E_i^{j_i^{-1}} \varDelta C_i^{j_i^{-1}}) = 0.$$

In a way similar to (5), it follows that

$$P_{\xi_1^n\eta}\left(\bigtimes_{i=1}^{n} E_i^{j_i} \times F^j\right) = P_{\zeta_1^n\eta}\left(\bigtimes_{i=1}^{n} C_i^{j_i} \times F^j\right). \tag{8}$$

In consequence of (7) and (8), there exists, for any $\bigtimes_{i=1}^{n} \delta_{X_i} \times \delta_Y$, a partition $\bigtimes_{i=1}^{n} \delta_{Z_i} \times \delta_Y$ such that

$$J\left(P_{\xi_1^n\eta}, Q_{\xi_1^n|\eta}; \bigtimes_{i=1}^{n} \delta_{X_i} \times \delta_Y\right) = J\left(P_{\zeta_1^n\eta}, Q_{\zeta_1^n|\eta}; \bigtimes_{i=1}^{n} \delta_{Z_i} \times \delta_Y\right). \tag{9}$$

Combining (9) with Proposition 2, we have the inequality (2).

The following corollaries are simple applications of Propositions 4.

COROLLARY 1. *If f is a B_{X_1}-measurable function, then*

$$I(\xi_1, \xi_2, \ldots, \xi_n \mid \eta) \geqslant I(f(\xi_1), \xi_2, \ldots, \xi_n \mid \eta).$$

COROLLARY 2. *We have*

$$I(\xi_1, \ldots, \xi_n \mid \eta) \geqslant I(\xi_1, \ldots, \xi_{n-1} \mid \eta).$$

4. Some Properties of Multivariate Information

We shall show in this section two theorems. Both are concerned with the decomposition of multivariate information. Theorem 1 describes that a multivariate information can be decomposed as a sum of a conditional multivariate information and certain mutual informations under some assumptions. The equality in the theorem may give one of the interpretations of conditional multivariate information.

Theorem 2 ensures a decomposition of multivariate information by clustering the random variables. That is, a multivariate information can be explained by a sum of the multivariate informations within the clusters and the multivariate information among the clusters, if adequate assumptions are satisfied.

Breafly speaking, Theorems 1 and 2 give us a decomposition by specifying one of the random variables and by clustering the random variables, respectively.

Now we show the following proposition, in a way similar to the well-known theorem which is concerned with integral representation of mutual information (Gelfand and Yaglom, 1957).

PROPOSITION 5. (a) *If* $P_{\xi_1^n} \ll Q_{\xi_1^n}$, *i.e., if* $P_{\xi_1^n}$ *is absolutely continuous with respect to* $Q_{\xi_1^n}$, *then*

$$I(\xi_1^n) = \int_{X_1^n} \log a(x_1^n)\, dP_{\xi_1^n},$$

where $a(x_1^n)$ *is the derivative of* $P_{\xi_1^n}$ *with respect to* $Q_{\xi_1^n}$.

(b) *If* $P_{\xi_1^n \eta} \ll Q_{\xi_1^n | \eta}$, *then*

$$I(\xi_1^n \mid \eta) = \int_{X_1^n \times Y} \log \bar{a}(x_1^n, y)\, dP_{\xi_1^n \eta},$$

where $\bar{a}(x_1^n, y)$ *is the derivative of* $P_{\xi_1^n \eta}$ *with respect to* $Q_{\xi_1^n | \eta}$.

THEOREM 1. *If* $P_{\xi_1^n \eta} \ll Q_{\xi_1^n | \eta}$, *then*

$$I(\xi_1, \ldots, \xi_n, \eta) = I(\xi_1, \ldots, \xi_n \mid \eta) + \sum_{i=1}^{n} I(\xi_i, \eta),$$

where $Q_{\xi_1^n \eta}$ *is the product measure on* $(X_1^n \times Y, B_{X_1^n \times Y})$, *which satisfies for any* $E_i \in B_{X_i}$, $i = 1, \ldots, n$, *and for any* $F \in B_Y$,

$$Q_{\xi_1^n \eta}\left(\bigtimes_{i=1}^{n} E_i \times F\right) = \prod_{i=1}^{n} P_{\xi_i}(E_i) \cdot P_\eta(F).$$

Proof. Let $b(x_1^n, y)$ be the derivative of $Q_{\xi_1^n | \eta}$ with respect to $Q_{\xi_1^n \eta}$. From the standard result of measure theory and our assumption, it follows that

$$a(x_1^n, y) = \bar{a}(x_1^n, y) \cdot b(x_1^n, y) \qquad \text{a.s.,} \tag{10}$$

where $a(x_1^n, y)$ is the derivative of $P_{\xi_1^n \eta}$ with resepct to $Q_{\xi_1^n \eta}$.

As we assume here that all conditional probabilities are measures, there exist the derivatives, $b_i(x_i, y)$, $i = 1, \ldots, n$, of $P_{\xi_i | \eta}(\cdot \mid y)$ with respect to $P_{\xi_i}(\cdot)$ for almost every y, because $P_{\xi_i | \eta}(\cdot \mid y)$ is absolutely continuous with respect to $P_{\xi_i}(\cdot)$. For any rectangle $\bigtimes_{i=1}^{n} E_i \times F$, $E_i \in B_{X_i}$, $F \in B_Y$,

$$Q_{\xi_1^n | \eta}\left(\bigtimes_{i=1}^{n} E_i \times F\right) = \int_F \prod_{i=1}^{n} P_{\xi_i | \eta}(E_i \mid y)\, dP_\eta$$

$$= \int_{\bigtimes_{i=1}^{n} E_i \times F} \prod_{i=1}^{n} b_i(x_i, y)\, dQ_{\xi_1^n \eta}. \tag{11}$$

According to the definition of $b(x_1{}^n, y)$ and (11), the following equation

$$\int_{\mathsf{X}_{i=1}^n E_i \times F} b(x_1{}^n, y)\, dQ_{\xi_1{}^n\eta} = \int_{\mathsf{X}_{i=1}^n E_i \times F} \prod_{i=1}^n b_i(x_i\,, y)\, dQ_{\xi_1{}^n\eta}$$

holds for any rectangle $\mathsf{X}_{i=1}^n E_i \times F \in \mathsf{X}_{i=1}^n B_{X_i} \times B_Y$.

Since $B_{X_1{}^n \times Y}$ is the σ-algebra generated by $\mathsf{X}_{i=1}^n B_{X_i} \times B_Y$, it follows that

$$b(x_1{}^n, y) = \prod_{i=1}^n b_i(x_i\,, y) \quad \text{a.s.} \tag{12}$$

Similarly, denoting by $a_i(x_i\,, y)$ the derivative of $P_{\xi_i\eta}$ with respect to $Q_{\xi_i\eta}$, we have

$$\begin{aligned} P_{\xi_i\eta}(E_i \times F) &= \int_F P_{\xi_i|\eta}(E_i \mid y)\, dP_\eta \\ &= \int_F \int_{E_i} b_i(x_i\,, y)\, dP_{\xi_i}\, dP_\eta = \int_{E_i \times F} b_i(x_i\,, y)\, dQ_{\xi_i\eta}\,, \end{aligned}$$

while

$$P_{\xi_i\eta}(E_i \times F) = \int_{E_i \times F} a_i(x_i\,, y)\, dQ_{\xi_i\eta}\,,$$

from which we easily have, for each i,

$$b_i(x_i\,, y) = a_i(x_i\,, y) \qquad \text{a.s.} \tag{13}$$

It should be noted here that $P_{\xi_1{}^n\eta} \ll Q_{\xi_1{}^n\eta}$ ensures $P_{\xi_i\eta} \ll Q_{\xi_i\eta}$.

In view of (10), (12), and (13), it is immediate that

$$a(x_1{}^n, y) = \bar{a}(x_1{}^n, y) \cdot \prod_{i=1}^n a_i(x_i\,, y) \quad \text{a.s.} \tag{14}$$

Our theorem follows from (14) and Proposition 5.

According to Theorem 1, we have

$$I(\xi_1\,,\ldots,\, \xi_{k-1}\,,\, \xi_{k+1}\,,\ldots,\, \xi_n \mid \xi_k) = I(\xi_1\,,\ldots,\, \xi_n) - \sum_{i \neq k} I(\xi_i\,,\, \xi_k).$$

Therefore, the conditional multivariate information among $\xi_1\,,\ldots,\, \xi_{k-1}\,,$ $\xi_{k+1}\,,\ldots,\, \xi_n$ relative to ξ_k may be intuitively interpreted as the remainder term when we subtract "the sum of the mutual informations between ξ_k and each of the other random variables" from the multivariate information of the system, $(\xi_1, ..., \xi_n)$. It follows that if the conditional multivariate information relative to some ξ_k is sufficiently small, then ξ_k may be considered as a regulating factor of all the system components. Moreover, we can find such a factor, if it exists, by computing the mutual informations. Because we have

$$\min_{1 \leqslant k \leqslant n} I(\xi_1, ..., \xi_{k-1}, \xi_{k+1}, ..., \xi_n \mid \xi_k) = I(\xi_1, ..., \xi_n) - \max_{1 \leqslant k \leqslant n} \sum_{i \neq k} I(\xi_i, \xi_k),$$

where the first term in the right side is constant for any choice of k.

THEOREM 2. *Let $Q_{(\xi_1^{n-1})\xi_n}$ be the measure satisfying $Q_{(\xi_1^{n-1})\xi_n}(\mathsf{X}_{i=1}^n E_i) = P_{\xi_1^{n-1}}(\mathsf{X}_{i=1}^n E_i) \cdot P_{\xi_n}(E_n)$ for each $\mathsf{X}_{i=1}^n E_i \in \mathsf{X}_{i=1}^n B_{X_i}$. If $P_{\xi_1^n} \ll Q_{(\xi_1^{n-1})\xi_n}$ and $P_{\xi_1^{n-1}} \ll Q_{\xi_1^{n-1}}$, then*

$$P_{\xi_1^n} \ll Q_{\xi_1^n},$$

and

$$I(\xi_1^n) = I((\xi_1, ..., \xi_{n-1}), \xi_n) + I(\xi_1^{n-1}), \tag{15}$$

where $I((\xi_1, ..., \xi_{n-1}), \xi_n) = I(\xi^{n-1}_1, \xi_n)$ is the mutual information between $(\xi_1, ..., \xi_{n-1}) = \xi^{n-1}_1$ and ξ_n.

Proof. From the assumption $P_{\xi_1^{n-1}} \ll Q_{\xi_1^{n-1}}$, there exists the derivative of $P_{\xi_1^{n-1}}$ with respect to $Q_{\xi_1^{n-1}}$, and we denote it by $a_{n-1}(x^{n-1}_1)$.

Using well-known results of integration theory, we obtain, for any $E^{n-1}_1 \in B_{X_1^{n-1}}$ and $E_n \in B_{X_n}$,

$$Q_{(\xi_1^{n-1})\xi_n}(E_1^{n-1} \times E_n) = \int_{E_1^{n-1} \times E_n} a_{n-1}(x_1^{n-1})\, dQ_{\xi_1^n}.$$

This equation can be extended to any $B_{X_1^n}$-measurable set E, i.e.,

$$Q_{(\xi_1^{n-1})\xi_n}(E) = \int_E a_{n-1}(x_1^{n-1})\, dQ_{\xi_1^n}. \tag{16}$$

Therefore,

$$Q_{(\xi_1^{n-1})\xi_n} \ll Q_{\xi_1^n}. \tag{17}$$

Combining (17) with the first assumption, we have

$$P_{\xi_1^n} \ll Q_{\xi_1^n}.$$

Let $a^*(x_1^n)$ denote the derivative of $P_{\xi_1^n}$ with respect to $Q_{(\xi_1^{n-1})\xi_n}$.

Then we have with $Q_{\xi_1{}^n}$-measure 1

$$a(x_1{}^n) = a^*(x_1{}^n) \cdot a_{n-1}(x_1^{n-1}), \tag{18}$$

because Eq. (16) implies that the derivative of $Q_{(\xi_1^{n-1})\xi_n}$ with respect to $Q_{\xi_1{}^n}$ must coincide with $a_{n-1}(x^n{}_1^{-1})$, with $Q_{\xi_1{}^n}$-measure 1.

According to (18) and Proposition 5, we obtain Eq. (15).

COROLLARY 1. *If, for some k,* $1 < k < n$,

$$P_{\xi_1{}^n} \ll Q_{(\xi_1{}^k)(\xi_{k+1}^n)}, \qquad P_{\xi_1{}^k} \ll Q_{\xi_1{}^k} \qquad and \qquad P_{\xi_{k+1}^n} \ll Q_{\xi_{k+1}^n},$$

then

$$I(\xi_1, \ldots, \xi_n) = I((\xi_1{}^k), (\xi_{k+1}^n)) + I(\xi_1{}^k) + I(\xi_{k+1}^n).$$

COROLLARY 2. *If* $P_{\xi_1{}^n} \ll Q_{(\xi_1^{n-1})\xi_n}, P_{\xi_1^{n-1}} \ll Q_{(\xi_1^{n-2})\xi_{n-1}}, \ldots, P_{\xi_1{}^2} \ll Q_{\xi_1{}^2}$, *then*

$$I(\xi_1{}^n) = \sum_{k=1}^{n-1} I((\xi_1{}^k), \xi_{k+1}).$$

Corollary 1 can be obtained in a way similar to Theorem 2. And we have Corollary 2 by repeated application of Theorem 2.

ACKNOWLEDGMENT

The author acknowledges the continuing guidance and encouragement of Professors H. Sakamoto and Y. Hayashi of Keio University. The author also thanks Y. Kato for his helpful advice.

RECEIVED: February 23, 1973; REVISED: August 29, 1973

REFERENCES

DOBRUSHIN, R. L. (1959), General formulation of Shannon's basic theorems of the theory of information, *Uspehi Mat. Nauk.* **14**, No. 6, 3–104.

GELFAND, I. M., AND YAGLOM, A. M. (1957), Calculation of the amount of information about a random function contained in another such function, *Uspehi Mat. Nauk.* **12**, No. 1, 3–52.

KOLMOGOROV, A. N., YAGLOM, A. M., AND GELFAND, I. M. (1956), Quantity of information and entropy for continuous distributions, Proceedings of the Third All-Union Mathematics Conference, Vol. 3, Izd. Akad. Nauk. SSSR.

KULLBACK, S., AND LEIBLER, R. A. (1951), Information and sufficiency, *Ann. Math. Statist.* 22, 79–86.

PEREZ, A. (1959), Information theory with abstract alphabets, *Theor. Probability Appl.* 4, 499–519.

PINSKER, M. S. (1960), "Information and information stability of random variables and processes," *Izd. Akad. Nauk. SSSR.* (Transl. Holden-Day, San Francisco, 1964.)

WATANABE, S. (1960), Information theoretical analysis of multivariate correlation, *IBM J. Res. Develop.* 4, 66–82.

WATANABE, S. (1969), "Knowing and Guessing: A Quantitative Study of Inference and Information," John Wiley and Sons, New York.